

UNA SERIE SULL'INTELLIGENZA ARTIFICIALE

Io, Gemini e altri quattro

Parte IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

ROCCO MASSIMILIANO MONDO

ABSTRACT

Sin dal primo episodio di questa serie, ho parlato di «democratizzazione tecnologica» come grande tema della rivoluzione IA. In questo nono episodio, quella promessa diventa progetto concreto. Il Progetto **ADAM (Adaptive Digital Assistant for Machine-learning)** è la sintesi di tutto ciò che abbiamo esplorato: **un sistema IA locale, modulare e specializzato**, progettato per la ricerca ingegneristica avanzata. Questo episodio ed il prossimo serviranno a delinearne il DNA: la filosofia progettuale, l'hardware target, l'architettura software e le motivazioni che mi hanno portato a scegliere un approccio completamente locale e open-source. In questo episodio ci occuperemo di definire la **parte software**, quindi realizzeremo una vera e propria mappa funzionale del modello.

PARTE IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

PREMESSA: DALLA SERIE AL PROGETTO

Siamo arrivati al penultimo episodio di questa serie. Guardando indietro, il percorso è stato notevole: dalla panoramica macroeconomica del CES 2026 (Parte I), alla storia dell'IA: da Turing ai robot neurali (Parte II), fino all'anatomia tecnica dei Transformer (Parte III), al prompt engineering (Parte IV), al confronto tra i cinque grandi LLM (Parte V), agli agenti IA (Parte VI), al RAG (Parte VII) e al fine-tuning per l'addestramento (Parte VIII).

Ogni episodio ha rappresentato un piccolo mattone. **Il progetto A.D.A.M.** è l'edificio.

Benché il nome possa rievocare il concetto biblico del primo uomo (o della prima umanità) - il che può andare bene visto che è la prima IA che sto creando - in effetti il nome ha un significato meno pragmatico e più tecnico: **A.D.A.M. è l'acronimo di Adaptive Digital Assistant for Machine-learning**. Il progetto nasce dall'ambizione di creare qualcosa che non sia un semplice **chatbot**, ma un sistema progettato per il potenziamento della ricerca ingegneristica in modo strutturale: un assistente che conosce il dominio tecnico perché è stato addestrato su di esso, che opera localmente perché la privacy dei dati professionali non è negoziabile, e che cresca nel tempo man mano che il dominio di conoscenza si espande. **A.D.A.M.** è chiamato ad operare nel campo dell'ingegneria elettrica ma con un **super obiettivo**: applicare le recenti scoperte fatte nel campo della **meccanica quantistica** per sviluppare nuovi e rivoluzionari prodotti, come ad esempio un prototipo di interruttore MT che usa la tecnica quantistica per l'interruzione dell'arco elettrico. Questo è stato sempre l'obiettivo principale, lo sviluppo di "modelli concettuali" (concept model) avanzati che si spingono fino alla ricerca di confine (**borderline**).

Con questa premessa si capisce adesso, il motivo per cui si ritiene necessario realizzare una IA specifica, con un DNA molto particolare.

In questo episodio verranno mostrati quali sono i principi progettuali che hanno portato alla realizzazione di **A.D.A.M.**, vedremo come è composta la vera anima di questa IA per poi definire una adeguata struttura fisica capace di ospitarla, alla fine avremo un vero ricercatore scientifico in ambito **STEM** (Science, Technology, Engineering and Mathematics) capace di collaborare attivamente allo sviluppo scientifico di progetti importanti.

PARTE IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

PERCHÉ UN PROGETTO PROPRIO?

La domanda è legittima: perché costruirsi un sistema IA quando ne esistono già di ottimi e disponibili con un abbonamento mensile? La risposta, che ho maturato nel corso di questi mesi di utilizzo parallelo di GPT, Gemini, Claude e LLaMA, è articolata in tre punti fondamentali.

1. Sovranità dei dati

Nel lavoro professionale di un progetto innovativo, i documenti tecnici rappresentano patrimonio intellettuale. Relazioni di calcolo, schemi di impianto, specifiche normative, documentazione di progetto — tutto ciò che viene condiviso con un **LLM cloud** entra in un ecosistema che non controlliamo. Un sistema locale elimina questa dipendenza alla radice: i dati non lasciano mai il proprio hardware.

2. Specializzazione verticale

I grandi modelli cloud sono generalisti per definizione. Sono eccellenti in compiti trasversali, ma quando si tratta di ragionare su normative CEI specifiche, calcoli illuminotecnici secondo UNI EN 12464-1 o verifica di selettività tra protezioni, le loro risposte sono spesso accurate nella forma ma imprecise nella sostanza. Un modello addestrato su **dataset** tecnici specifici, anche se più piccolo, batte sistematicamente un modello generalista su compiti di dominio ristretto.

3. Autonomia e controllo

Costruendo un sistema proprietario possiamo controllare ogni variabile: la versione del modello, i parametri di inferenza, le restrizioni etiche, il comportamento in caso di incertezza. Non dipenderà da aggiornamenti non annunciati che cambiano il comportamento del modello stesso o da politiche di **content policy** che possono variare o peggio da interruzioni di servizio cloud.

Esiste anche un'altra componente che mi ha spinto verso questa strada ed è quello della ricerca, la propensione a spingersi un po' oltre per guardare il futuro che verrà e che altri potranno vedere magari fra un decennio se non di più.

PARTE IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

L'IDEA DI BASE SU CUI SI FONDA IL PROGETTO

L'idea principale su cui si fonda il **progetto A.D.A.M.** è quella di riprodurre in qualche modo il funzionamento del cervello umano relativamente alla gestione della memoria a lungo e breve termine e alla capacità di sviluppare una risposta autonoma basata sia su una **conoscenza strutturata ma generalizzata in ambito STEM** (DNA STEM) che su una **conoscenza specifica** realizzata attraverso un meccanismo **RAG** molto sofisticato che interviene sulla capacità di A.D.A.M. di trattare argomenti specifici come la meccanica quantistica **di R. Feynman** applicata al campo dell'ingegneria elettrica e nello stesso tempo di apprendere dal lavoro che viene svolto durante lo sviluppo dello progetto stesso (**extended context**). Molti di voi avranno avuto modo di conoscere la teoria che vede il cervello umano diviso in **due emisferi**, destro e sinistro, collegati dal **corpo calloso**, che lavorano in sinergia per ogni attività. **L'emisfero sinistro** controlla la parte destra del corpo ed è specializzato nella **logica, nel linguaggio, nella matematica e nel pensiero sequenziale**. L'emisfero destro controlla la parte sinistra del corpo ed eccelle nella **creatività, nella percezione spaziale, nell'intuizione e nell'elaborazione delle emozioni**. Quindi ho pensato di dividere il cervello di A.D.A.M. in **due parti**, quello **sinistro** legato alla logica STEM e quello **destro** legato all'assimilazioni di nuovi concetti attraverso il meccanismo RAG e l'integrazione del **"contesto esteso"**, ossia l'inclusione contestuale dei progetti nel campo del dominio di nostro interesse.

Con questa idea ho iniziato a strutturare **A.D.A.M.** seguendo una linea ben precisa, stabilendo anche un carattere di base **"STEM oriented"** e una propensione adattiva allo studio di problematiche legate alla fisica quantistica.

In questo senso possiamo rappresentare il cervello di A.D.A.M. come formato da due parti:

- **ADAM Core**
- **ADAM Plus**

Effettivamente ADAM è composto fisicamente da due parti distinti, una che implementa il carattere STEM (ADAM Core) e l'altra parte che implementa la conoscenza specifica nel campo di nostro interesse attraverso la capacità di crescere con i risultati ottenuti durante lo studio degli stessi progetti e l'implementazione di nuove nozioni specifiche in campo ingegneristico.

PARTE IX

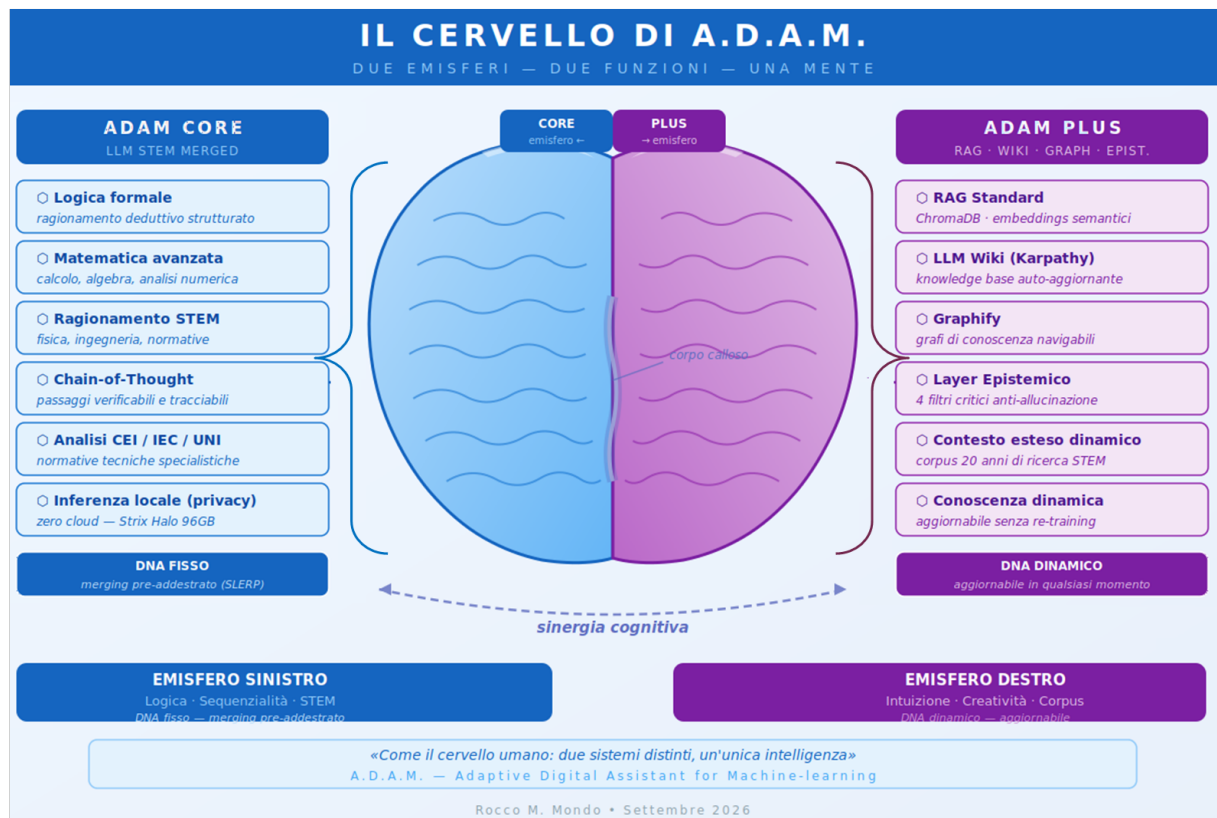
Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

Questa soluzione ci permette di superare un grosso limite degli attuali LLM standard e cioè quello della staticità dell'informazione di cui abbiamo parlato nell'episodio VII dedicato al RAG, dove abbiamo visto che ogni LLM ha una data di emissione che determina l'invecchiamento dell'informazione contenuta.

D'altronde pensare di addestrare da zero una IA ogni volta che vogliamo ampliare la sua conoscenza è una strada impercorribile, perché i costi da sostenere, in termini economici e di risorse più in generale sono alla portata solo dei gradi big come NVideo, Anthropic o Google solo per citarne alcuni.

Quindi la soluzione adottata è quella di avvalersi di una **LLM** con base **STEM** avente una formamentis tecnica, non legata fondamentalmente all'aggiornamento di dati, ma già addestrata da aziende specializzate nel ragionamento tecnico-scientifico e che sia open source (open weights) e un'altra parte che contenga un **corpus dati** (aggiornabile) orientato all'ingegneria elettrica e alla meccanica quantistica. Queste sono state le principali linee guida che ho seguito per implementare il progetto **A.D.A.M.**

5



PARTE IX

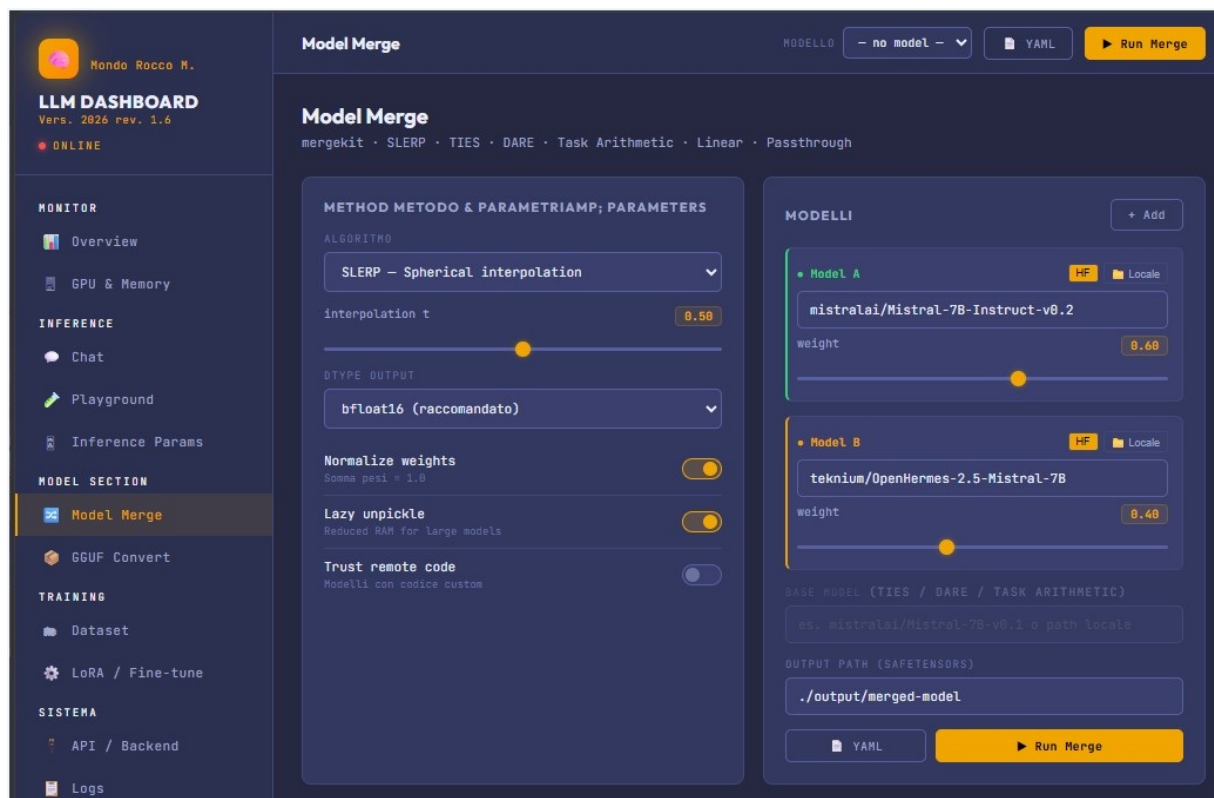
Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

COME HO IMPLEMENTATO A.D.A.M. CORE

Come già specificato nel capitolo precedente, implementare una LLM da zero sarebbe un'impresa veramente titanica, sia da un punto di vista delle risorse umane che da quelle economiche. Quindi ho preso la decisione di adottarne una che rispondesse ai requisiti di progetto che sono stati trattati nel paragrafo precedente. Volendo essere sicuro di avere un motore STEM molto efficiente ho pensato di non accontentarmi di un solo LLM ma di fonderne due insieme attraverso un processo di **"merging"**, un processo che unisce due LLM compatibili per amplificarne le capacità. Benché questo processo non richieda delle particolari conoscenze ingegneristiche, la ricerca di due DNA compatibili ha richiesto molto impegno e tempo, ma dopo qualche mese di ripetuti test sono approdato ad una soluzione.

Ottenuto il nuovo LLM, ho voluto ottimizzarlo ulteriormente attraverso un processo di fine-tuning leggero, ottenendo così **A.D.A.M. Core**.

Per velocizzare i processi di **merging**, **fine-tuning** e **cycle-testing** ho dovuto creare un programma che lavorasse per me ogni giorno e tutti i giorni, quindi ho creato una **dashboard** che mi permettesse di controllare i processi e i loro parametri. Questo ha velocizzato notevolmente la **pipeline** generale legata al processo di reazione e addestramento.



LLM Dashboard Studio

PARTE IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

Non sono entrato nel dettaglio delle specifiche operazioni, perché sarebbe stato veramente impegnativo trascrivere il lavoro fatto, ma ho cercato di sintetizzare in poche righe mesi di lavoro. Quello che più mi preme è illustrare i principi su cui ho basato l'intero progetto così che possano essere un utile riferimento per chi voglia cimentarsi in questo tipo di attività.

Prima di passare alla fase di realizzazione di **A.D.A.M. Plus**, vorrei spendere alcune parole sull'argomento hardware, ossia la parte fisica che ospita l'intero sistema IA.

Le scelte e l'implementazione di questo sistema è stato ed è fortemente influenzato dal tipo di hardware utilizzato (vedremo i dettagli nel prossimo episodio), di fatto avere la disponibilità di un hardware potente e scalabile significa poter implementare un sistema IA più potente e flessibile.

Essendo un ricercatore indipendente, non sovvenzionato, ho dovuto affrontare autonomamente tutte le spese legate a questo progetto e quindi ho dovuto fare delle scelte non solo tecniche ma anche legate all'aspetto economico.

Nel prossimo episodio parleremo proprio del bilanciamento tecnico-economico che poi ha portato alla scelta della piattaforma su cui "vive" il **sistema A.D.A.M.**

IL DOMINIO DI CONOSCENZA: A.D.A.M. PLUS

Un sistema IA specializzato vale esattamente quanto il dominio di conoscenza che gli viene fornito. Per il progetto ADAM, il dominio tecnico di riferimento è l'ingegneria elettrica e la meccanica quantistica. Questa è la base su cui ho costruito e sto costruendo **A.D.A.M. PLUS**.

Definito il dominio di interesse ho concentrato tutta la mia attività nel realizzare un **sistema RAG** che mi permettesse di **integrare, mantenere ed ampliare** questo corpus di conoscenza all'interno del sistema. L'applicazione di un semplice meccanismo RAG non si è dimostrato sufficiente per garantire la crescita del sistema stesso, occorre qualcosa di più sofisticato e più orientato alla manipolazione e gestione della conoscenza. Per ottenere questo risultato ho dovuto adottare il pattern **"LLM Wiki"** proposto da **Andrej Karpathy** per superare il limite fondamentale della **RAG** (Retrieval-Augmented Generation). Il pattern di **Karpathy** mi ha permesso non solo di strutturare

8

Rocco M. Mondo, Settembre 2026

PARTE IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica

Adesso l'idea del progetto è completa e dovrebbe essere più accessibile: un sistema composto e non monolitico, con parti ampliabili e sostituibili, scalabile nel tempo a seconda delle esigenze.

I FILTRI EPISTEMICI

Dopo mesi di lavoro sul modello base di A.D.A.M. (rudimentale e ancora poco istruito) ho riscontrato il tipico comportamento servile delle IA, cioè la tendenza di dare sempre una risposta ragionevole e positiva, mettendo a suo agio l'interlocutore, quasi con riverenza.

Questo passaggio è sempre evidente durante una comunissima inferenza, ma per il compito assegnato ad A.D.A.M. questo non va assolutamente bene. In un contesto di ricerca di confine, bisogna trovare e studiare soluzioni reali e soprattutto dubitare delle stesse soluzioni. Questo non accadeva all'interno dei test preliminari svolti su A.D.A.M., perché il suo nucleo deriva da una IA STEM addestrata con normali tecniche standard, quindi soggetta non solo ad "allucinazioni" ma anche alla mancanza di obiettività su ragionamenti che sono molto astratti. Questo problema non è di secondaria importanza per una IA che ha come obiettivo quello di essere un ricercatore scientifico nel campo ingegneristico/quantistico. La soluzione è arrivata dopo innumerevoli tentativi: ho applicato ad **ADAM PLUS** uno **strato epistemico** che mi ha permesso di aumentare la sua capacità di essere obiettivo. La soluzione (ancora sotto test) di applicare un "**layer epistemico**" ad ADAM PLUS di modo tale da modificare il suo comportamento deduttivo e proattivo nei confronti di argomenti astratti e di difficile verifica è stato vincente, almeno per il momento.

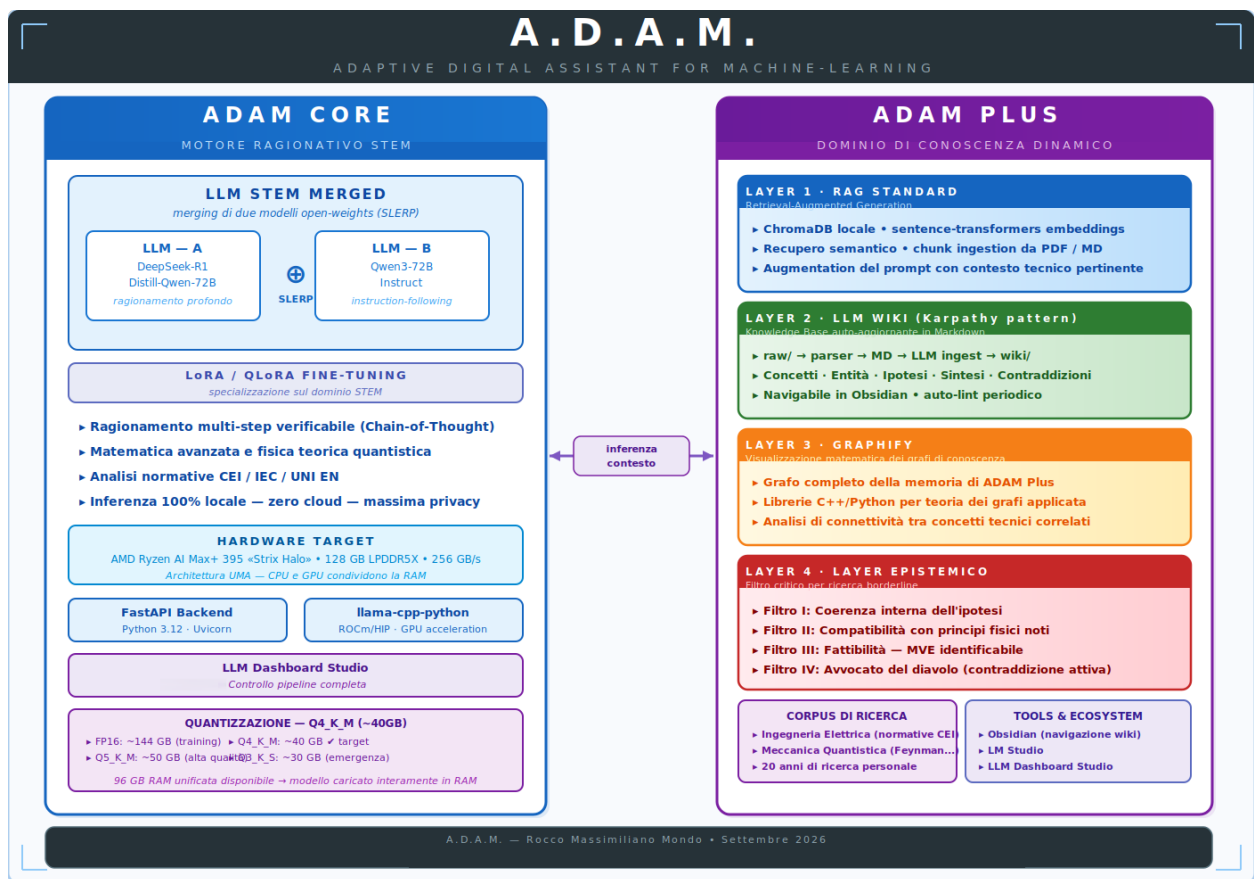
[Layer epistemico] – Direttamente da Wikipedia

Un layer epistemico (livello epistemico) è un concetto filosofico e cognitivo che riguarda l'architettura della conoscenza. Rappresenta il "filtro" o il "piano di realtà" attraverso cui interpretiamo, organizziamo e classifichiamo le informazioni (credenze, dati, verità), strutturando il modo in cui comprendiamo il mondo o un determinato sistema.

Per concludere vi lascio con una immagine che, in qualche modo, cerca di rappresentare **A.D.A.M.** nel suo complesso.

PARTE IX

Il Progetto A.D.A.M.: Un'IA Personale per la Ricerca Ingegneristica



CONCLUSIONE

Un'anima non può esistere sul piano materiale se non ha anche un corpo, quindi se vogliamo che ADAM diventi operativo nella quotidianità dobbiamo dargli anche un corpo. Nel prossimo episodio progetteremo un corpo per ADAM, capace di crescere nel tempo (scalabilità) così come farebbe un bambino nel diventare adulto.